# INVESTIGATION OF TRIPLE PLAY SERVICES

**Valentin Hristov, Mario Kanchev**

*South- West University – Blagoevgrad,*
*66 Ivan Mihajlov Str, Blagoevgrad, Bulgaria*
*Phone (+359)738889132; E-mail* v_hristov@aix.swu.bg .

*Abstract*

*Today there are three main mass communication services- telephony, television and Internet (conditionally, data). In past they were divided technologically and organizing: each was base on the own infrastructure/facilities, accordingly, different operators. For the user this looked as a telephone wire, television cable and Internet wire or wall-plug. Today, multiservice network's operator, which provides its users broadband IP connection (with speed not less than several megabits per second) can provide simultaneously these three the most mass and accustomed services through IP. This way of granting the services is named Triple Play.*

*The present paper aims at creating model, which simulates processes in the system, build on the principle of Triple Play and getting statistics about: Blocking probability of voice requests; Average time spent in the queueing system of video requests; Average time spent in the queueing system of data request and etc.*

## 1. INTRODUCTION

An important challenge in triple play networks is satisfying the requirements of different types of traffic (voice, video and data) within a single network. Real-time traffic e.g., voice and video require low, predictable delay in transmitting information from one location to another (end-to-end) or the result may be a distorted voice or videos received at the destination. On the other hand, elastic traffic e.g., data transmissions, e-mail, and file transfers not as sensitive to timing constraints, but are usually very sensitive to lost data. In earlier work [3] we find that packet losses in network nodes are very different depending on the ratio of different types of traffic sharing capacity of the network connections.

The problem of capacity sharing in the network was earlier researched for narrowband networks in terms of homogeneous utilizations (See [1] and its references). While VoIP telephony has been extensively studied in [6] and [7], blocking probability and setup time for telephony calls have received relatively less attention. The blocking probability and setup time has a direct impact on the users' satisfaction: The user is used to waiting a maximum of eleven seconds as specified in [8] and expects to experience the same even if the technology is different. Therefore, the blocking probability for telephony calls as well as average time spent in the queueing systems of video and data requests should be estimated and optimized. This is the motivation behind this article.

## 2. MODEL

In triple play networks the capacity of network is shared between voice, video (video on demands- VoD) and data. Under voice requests are implied telephony calls. Video request present itself inquiries for viewing a television (IPTV) or concrete videomaterial (VoD). Data requests are considered the data from applications, e.g. data transmissions, e-mail, and file transfers (encapsulated in IP packets), which are sent through Internet connection.

For processing these three type requests (named servers) some resources should be used. For each traffic type are dedicated a certain number of servers, i.e. voice server/s, video server/s and data server/s.

On fig. 1 corresponding Q – scheme is depicted. Each voice, video or data channel is approximated as a input buffer (Q1, Q2 or Q3), and server/s (VoS, ViS, and DS) for voice, video and data. The implementation in corresponding

General Purpose Simulation System (GPSS) model is based on the algorithm proposed in [4].

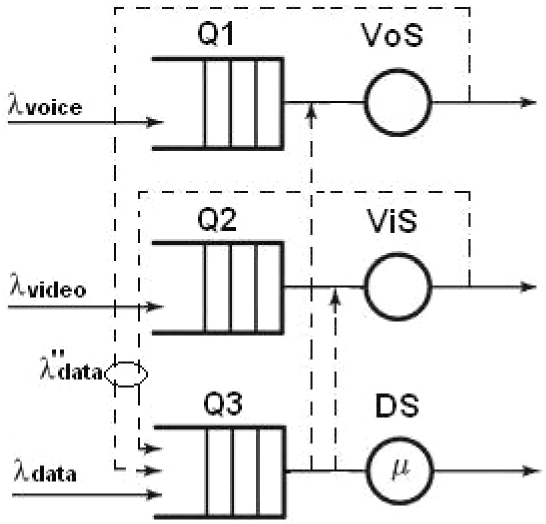


Fig. 1

As can see on fig.1 there are three types servicing in queuing systems which can be classified as follows:

- Priority service for voice and video requests in corresponding servers. The selection of a packet from the voice/video or data queue depends on priorities that are permanently assigned to these packets;
- First come, first served is realized by data server for data requests. Data packets are served by data server in the order they arrive;
- Preemption for data requests in voice and video servers. The data packet currently being serviced can be interrupted and preempted if a higher priority packet arrives in the queue. This is secondary data request (fig. 1).

As mentioned above maximum of session setup time [8] for telephony call (VoIP) is 11 seconds if triple play technology is used. Therefore, the principle of using the voice server(s) for processing the voice requests is following: new voice request is queued if all voice servers are occupied by processing the voice requests (received earlier); Every time when a voice request/transaction departs the queue and tries to enter in voice server, its time spent in the queueing system is checked; If it is less than 11 sec. the voice request enter in this server; Otherwise, this voice request is reported as blocked.

Video request enter in the queue (buffer) when all video servers are occupied by processing the video requests. Data requests can be serviced by data- , voice- and video- servers. If any voice and video server is free, the data request will get up on service. However, this data request will be displace from there in case of arrival voice or video request, corresponding (voice and video request have high priority).

The intervals between arrivals of (data, voice and video) requests are assumed with the exponential law of distribution, as enough realistic. The distribution of servicing times for the all three type requests is assumed with the exponential law, too.

Each one type of requests (data, voice and video) has different average interval between arrivals of requests or intensity, correspondingly, as well as different mean time of servicing.

## 2.1. Constants and parameters

For voice requests the mean time of servicing by voice server is $1/\mu_{VOICE}$ = 98 seconds. This is average length of occupation of individual phone line in hour of the most load [2]. The mean time of servicing of video requests is $1/\mu_{VIDEO}$ = 13600 seconds (TNS Gallup Media affirms that people looks a television 227 minutes per day). The mean time of data requests is $1/\mu_{DATA}$ = 100 seconds (this value was received as a result of statistical investigation of traffic from Bulgarian news sites).

The other parameter in model is servers' utilization (loads), $\rho = \lambda / \mu$ or the intensity of input flows or correspond to average interval between arrivals of requests. In order to achieve desired load in the model following formula is used to calculate average interval between arrivals of requests (transactions):

(1) $$\rho = \lambda / \mu \Leftrightarrow 1/\lambda = \rho / \mu$$

where

$\mu$ - mean time of servicing of corresponding type request;

$\rho$ - value of initial load, is defined on formula;

$\lambda$ - intensity of input flow requests, or $1/\lambda$ - average interval between transactions' arrivals.

In model is used a certain constants for number and states of servers: The amount of servers dedicated for voice, video and data requests are described with Num_of_VoS, Num_of_ViS, and Num_of_DS; The states of servers are described with following constants: 0 - is free, 1 - is occupy by processing a voice request, 2 - is occupy by processing a video request, 3 - is occupy by a data request.

In model a transaction parameter identifies type of the request: 1 - a voice, 2 - video, 3 - data. Another transaction parameter (at the secondary data requests)- P$residual_time is used for writing the time that remains to be service this transaction corresponding to data request.

### 2.2. Data requests

After data transaction is generated, the modeling system is examined for presence of free servers starting with data servers: If amongst them there is free, the request will be serviced by the found free data server. If there is not, the voice and video servers are examined. If after checking all servers free server was not found, the request enters in the queue and waits, while a server will not be free. After the completion of processing the request is removed from the system. Note, the role of secondary data requests and P$residual_time parameter when processing such request: If the value of this parameter is more than zero, this signifies that this request/ transaction has been displaced (while it is serviced by a voice or video server/s) and its remaining time of servicing is this value.

### 2.3. Voice requests

After voice transaction is generated, the system is examined for free voice server/s. If amongst them there is free, the request will be serviced by the found free server. If there is not free server, the system is examined for voice server which services a data transaction. If there is at least one server satisfying this condition, the processing of this data request is displace (it is moved in the queue) and voice transactions enters to be serviced in already freed server. The voice request is reported as blocked, if this waiting for server more than 11 seconds (and its corresponding GPSS transaction is terminated).

### 2.4. Video requests

Processing of the video requests is similarly as processing of voice requests. The only one exception is when free video server absences the request is not blocked (corresponding to this request transaction enters in the queue).

The statistical results received by modeling are written in the CSV-file which format is compatible with MS Excel.

## 3. VERIFICATION AND SIMULATION RESULTS

Verification is made as results from simulation model for three voice servers (Simulation of VoS and ViS servers is similar but more simple) are compared with these for M/M/a/a system (a=3).

| Ro | Erlang-B | Pb | Error , % |
|---|---|---|---|
| 0 | - | 0 | 0 |
| 0,1 | 0,003334568 | 0,00333 | 0,13 |
| 0,2 | 0,019823789 | 0,02098 | 5,83 |
| 0,3 | 0,05007212 | 0,0494 | 1,34 |
| 0,4 | 0,089775561 | 0,08957 | 0,23 |
| 0,5 | 0,134328358 | 0,132746 | 1,18 |
| 0,6 | 0,180267062 | 0,18091 | 0,36 |
| 0,7 | 0,22537782 | 0,224228 | 0,51 |
| 0,8 | 0,268406337 | 0,270692 | 0,85 |
| 0,9 | 0,308738412 | 0,311504 | 0,9 |
| 1 | 0,346153846 | 0,344573 | 0,46 |

Table 1

Bellow, for purposes of the verification, estimations of blocking probability are compared with results for blocking probability computed by the Erlang-B formula, as well as corresponding table is fill. Erlang-B formula calculates the probability that an arriving packet finds all servers busy but no waiting line for packets in the M/M/a/a system, or queueless model:

$$(2)\quad P\left[k(t)=a\right]=\rho_a=\frac{\rho_1^a}{a!}\rho_0=\frac{\rho_1^a}{a!\left(\sum_{i=0}^{a}\frac{\rho_1^i}{i!}\right)}$$

where: $p_0$- the probability of the first state; $a$ – number of servers; $\rho_1$ - load of each server in the M/M/a/a system ($\rho_1 = a.\rho$).

The difference between the blocking probability and its estimation is shown in last column in tabl.1. Unlike the difference for smaller values of utilization, e.g. $\rho$=0.2 which is 5.83%, the difference generally is less than 1.4%. So, because the GPSS model for smaller values of utilization is not pure M/M/3/3 system (there are significant number of voice request with queueing time less than 11 seconds, which are serviced, i.e. not blocked), it is useful to verify it with Erlang-C formula. This formula calculates the probability that an arriving packet finds all servers busy in a M/M/a system. This probability is the same as the probability that the waiting time for the packet in the queue is greater than zero, or the probability that the number of packets in the queueing system is greater than or equal to a=3:

(3) $P\,[K(t) \geq a] = p_a/(1-\rho)$ , where

$p_a$- the probability that all *a* servers are in use:

$$p_a = \frac{\rho_1^a}{a!} p_0$$

where: $p_0$- the probability of the first state. For the M/M/3 discipline, $\rho = \lambda/a\mu = 0.1$ (a = 3 servers are used, thus, $\rho_1 = \lambda/\mu = 0.3$):

$$(4)\quad P_0 = \frac{1}{\sum_{i=0}^{a-1} \frac{\rho_1^i}{i!} + \frac{\rho_1^i}{a!} \frac{1}{1-\rho}} = \frac{1}{\sum_{i=0}^{2} \frac{\rho_1^i}{i!} + \frac{\rho_1^3}{3!} \frac{1}{1-\rho}} = 0,7407407$$

Consecutively, substituting in (3) $p_a$=0,0033333, and $P\,[\,K(t) \geq 3\,] = p_a / (1 - \rho)$ = =0,003333/0.9=0,0037037.

The estimation of blocking probability (from simulation) is Pb= 0,0033333. The difference between the blocking probability and its estimation is less than 10%, which is acceptable for engineering purposes. Increasing system load leads to higher difference between these values, e.g. for $\rho$=0.3 ($\rho_1$=0.9) the blocking probability computed by the Erlang-C formula is $P\,[K(t) \geq 3]$= 0,0700288, but the blocking probability from simulation is Pb=0,0494. Thus, the difference is 29%, which is acceptable keeping in mind that (about simulation) a voice requests is blocked if its queuing time is greater than 11 sec. and we can conclude that proposed model is adequate [1].

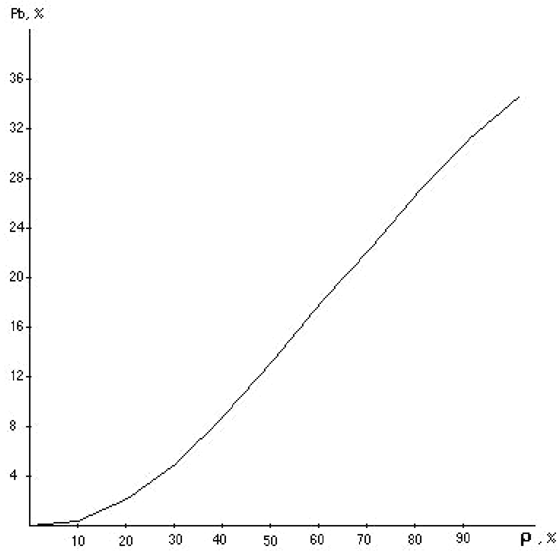


Fig. 2

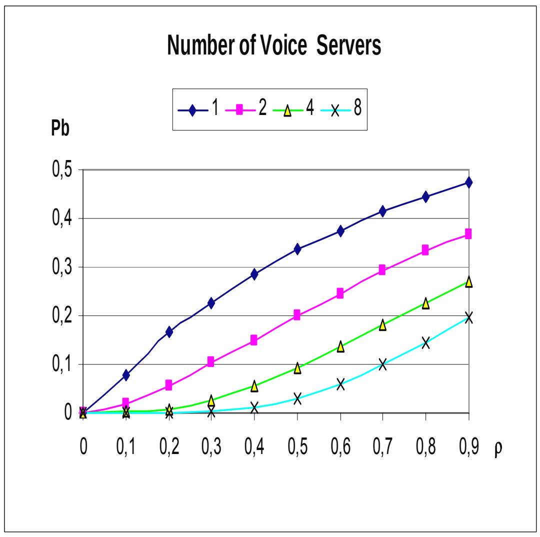


Fig. 3

On fig. 2 is depicted the blocking probability of voice requests- Pb vs. load of each voice server for three voice servers. Increasing the system load in voice channel leads to higher blocking probability of voice requests.

On fig. 3 is depicted the blocking probability of voice requests vs. load of each voice server for different number of voice servers. Increasing the number of voice servers for all system loads leads to lower blocking probability of voice requests (fig. 3).

Analogically, increasing the number of video servers leads to lower average queueing time of video requests- E[Tq] (fig. 4).

The results from figures 3 and 4 are logical due to statistical multiplexing at processing of voice/video requests from voice/video servers

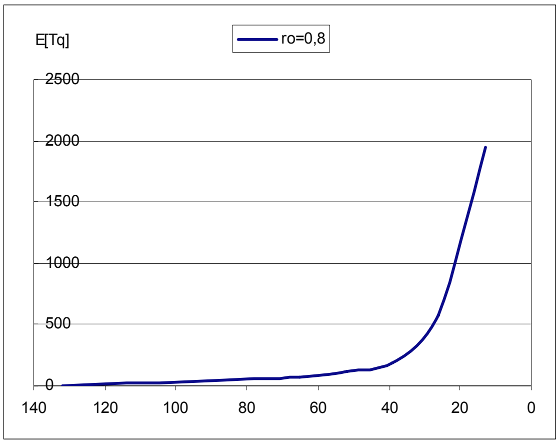


Fig. 4

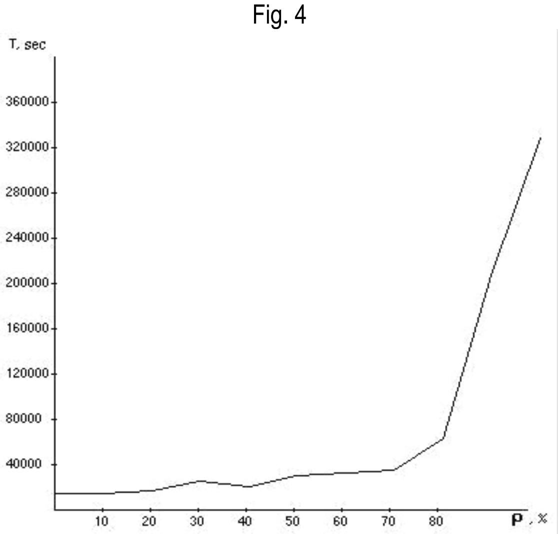


Fig. 5

On fig. 5 is depicted the average time spent in the queueing system of video requests vs. load of video server (the number of video servers is fixed to one). Increasing system loads in video channel leads to higher average time spent in the queueing system of video requests (fig. 5).

## 4. CONCLUSION

As this example shows, the proposed model gives us the opportunity of quickly comparing different situations corresponding to different values of the system parameters, for engineering purposes.

Keeping in mind simulation results depicted on fig. 2 and 3, we can conclude that it is useful to increase the number of voice servers in triple play networks.

## References

[1] Сурихин П. Л., Пономарев Д. Ю, Исследование вероятностно-временнных характеристик однофазных систем массового обслуживания с ограниченной очередью с помощью имитационного моделирования // Сборник научных трудов. Красноярск ИПФ – 2003., с. 420-425.

[2] Шелухин О.И. и др. Фрактальные процессы в телекоммуникациях. - М.: Радиотехника, 2003.

[3] Hristov V., et al. Efficient Sharing of the Capacity for Different Types Traffic in Multiservice Networks. Proc. of the Conference ELEKTRONIKA'-2008, Sofia, ISBN 954-90209-3-2

[4] Hristov V. , GPSS Simulator of Multiservice Networks, on-line available on: ice.prohosting.com/vhristov/smn.gps

[5] Panagiev, O. B. Nonlinear systems modeling in broadband communications. ICEST, Proc. of Papers, vol.1, Ohrid, 24-27 June 2007, pp. 321-324

[6] B. Goode, "Voice over Internet Protocol (VoIP)," Proc. IEEE, vol. 90, no. 9, 2002.

[7] A. Valko, A. Racz, and G. Fodor, "VoIP QoS in Third-Generation Mobile Systems," IEEE J. Selected Areas in Comm., vol. 17, no. 1, Jan. 1999.

[8] T.S.S. of ITU, "ITU-T Recommendation E. 721—Network Grade of Service Parameters and Target Values for Circuit-Switched Services in the Evolving ISDN," 1991.